

April 18<sup>th</sup>, 2018

## Mining the Integrated Connectedness of Biomedical Systems

*This seminar is part of the "Bioinfo4Women" series.*

**Natasa Przulj**

University College London (UCL)

### Abstract

We are faced with a flood of molecular and clinical data. Various bio-molecules interact in a cell to perform biological function, forming large, complex systems. Large-scale patient-specific omics datasets are increasingly becoming available, providing heterogeneous, but complementary information about cells, tissues and diseases. The challenge is how to mine these interacting, complex, complementary data systems to answer fundamental biological and medical questions. Dealing with them is nontrivial, because many questions we ask to answer from them fall into the category of



computationally intractable problems, necessitating the development of heuristic methods for finding approximate solutions.

We develop methods for extracting new biomedical knowledge from the wiring patterns of systems-level, heterogeneous, networked biomedical data. Our methods link the patterns in molecular networks and the multi-scale network organization with biological

function. In this way, we translate the information hidden in the wiring patterns into domain-specific knowledge. In addition, we introduce a versatile data fusion (integration) framework that can effectively integrate the information obtained from mining molecular networks with patient-specific somatic mutation data and drug chemical data to address key challenges in precision medicine: stratification of patients, prediction of driver genes in cancer, and re-purposing of approved drugs to particular patients and patient groups. Our new methods stem from novel network science approaches coupled with graph-regularized non-negative matrix tri-factorization, a machine learning technique for dimensionality reduction and co-clustering of heterogeneous datasets. We utilize our new framework to develop methodologies for performing other related tasks, including disease re-classification from modern, heterogeneous molecular level data, inferring new Gene Ontology relationships, and aligning multiple molecular networks.

## Short bio



Natasa Przulj is a Professor of Biomedical Data Science at University College London (UCL) Computer Science Department. She was elected into Academia Europaea, The Academy of Europe, in 2017. She is a Fellow of the British Computer Society. In 2014, she was awarded the British Computer Society Roger Needham Award for a distinguished research contribution in computer science by a UK based researcher within ten years of their PhD. She received two prestigious European Research Council (ERC) grants, the ERC Consolidator grant titled "Integrated Connectedness for a New Representation of Biology" (2018-2023) and the ERC

Starting Independent Researcher Grant titled "Biological Network Topology Complements Genome as a Source of Biological Information" (2012-2017). She held a prestigious NSF CAREER Award for the project titled "Tools for Analyzing, Modeling, and Comparing Protein-Protein Interaction Networks" in 2007-2011 at University of California Irvine. Her research has also been supported by other large governmental and industrial grants including those from GlaxoSmithKline, IBM and Google. She was previously an Associate Professor (Reader; 2012-2016) and Assistant Professor (Lecturer; 2009-2012) in the Department of Computing at Imperial College London and an Assistant Professor in the Computer Science Department at University of California Irvine (2005-2009). She obtained a PhD in Computer Science from University of Toronto in 2005.